

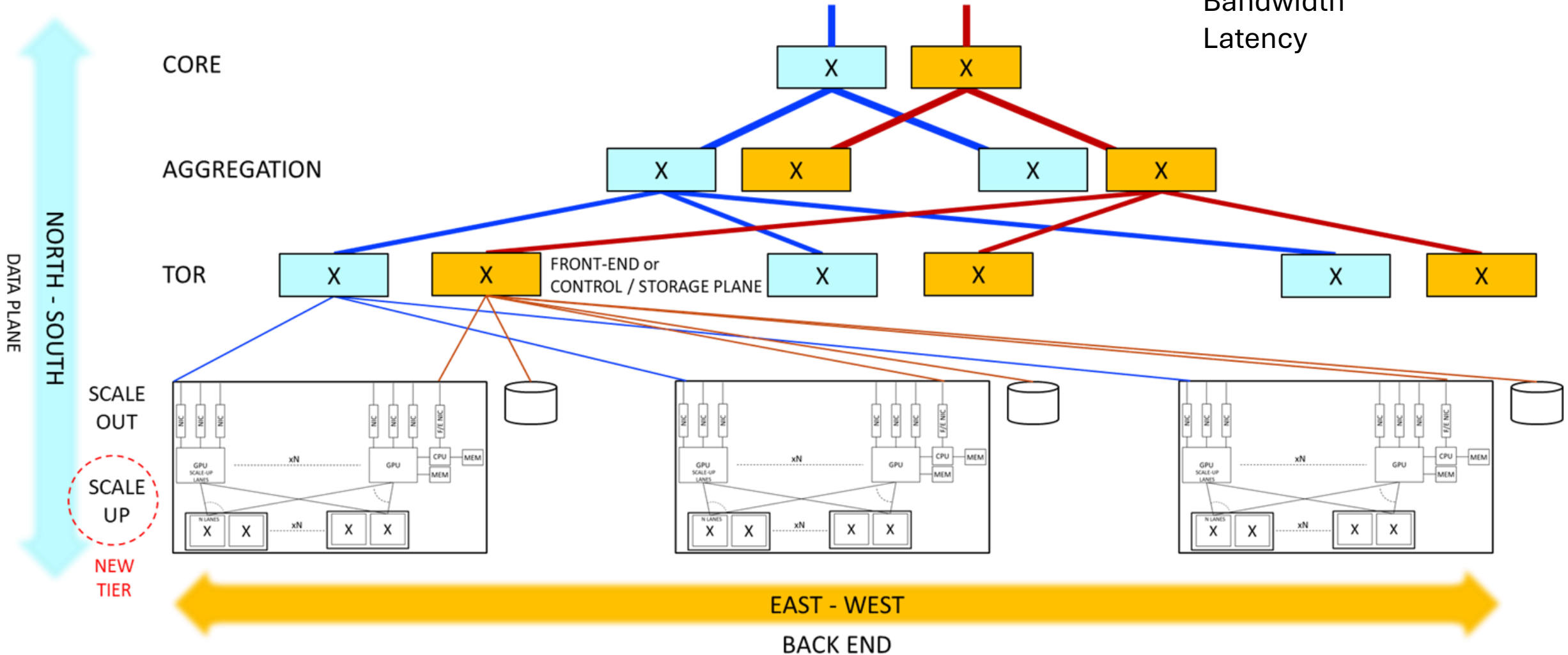
Scaling Fabric Technologies for AI Clusters

Darrin Vallis
Solution Architect, AMD

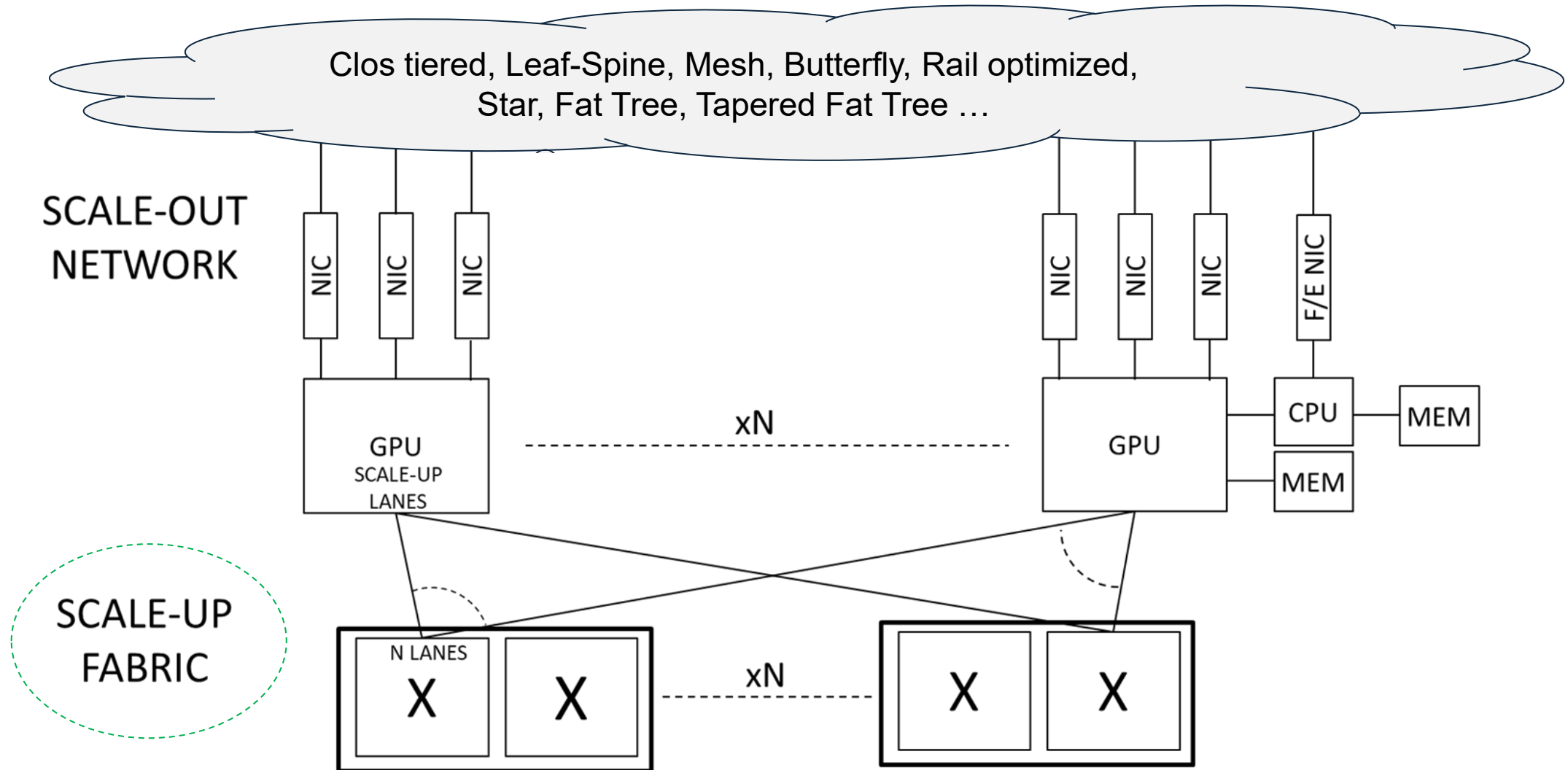
Scale-Up In Data Center Context

Network Tiers

Bandwidth
Latency



Scale-Up vs. Scale-Out



Scale-Up Fabric Technologies

NVLink¹



NVLink Fusion²



Photonic Fabric^{TM 3}



SUE⁴ - Scale Up Ethernet



UAL⁵ - Ultra Accelerator Link



1. <https://www.nvidia.com/en-us/data-center/nvlink/>

2. <https://www.nvidia.com/en-us/data-center/nvlink-fusion/>

3. https://www.kisacoresearch.com/sites/default/files/presentations/preet_virk_-_celestial_ai_-_photonic_fabRICTM_based_scale-up_network.pdf

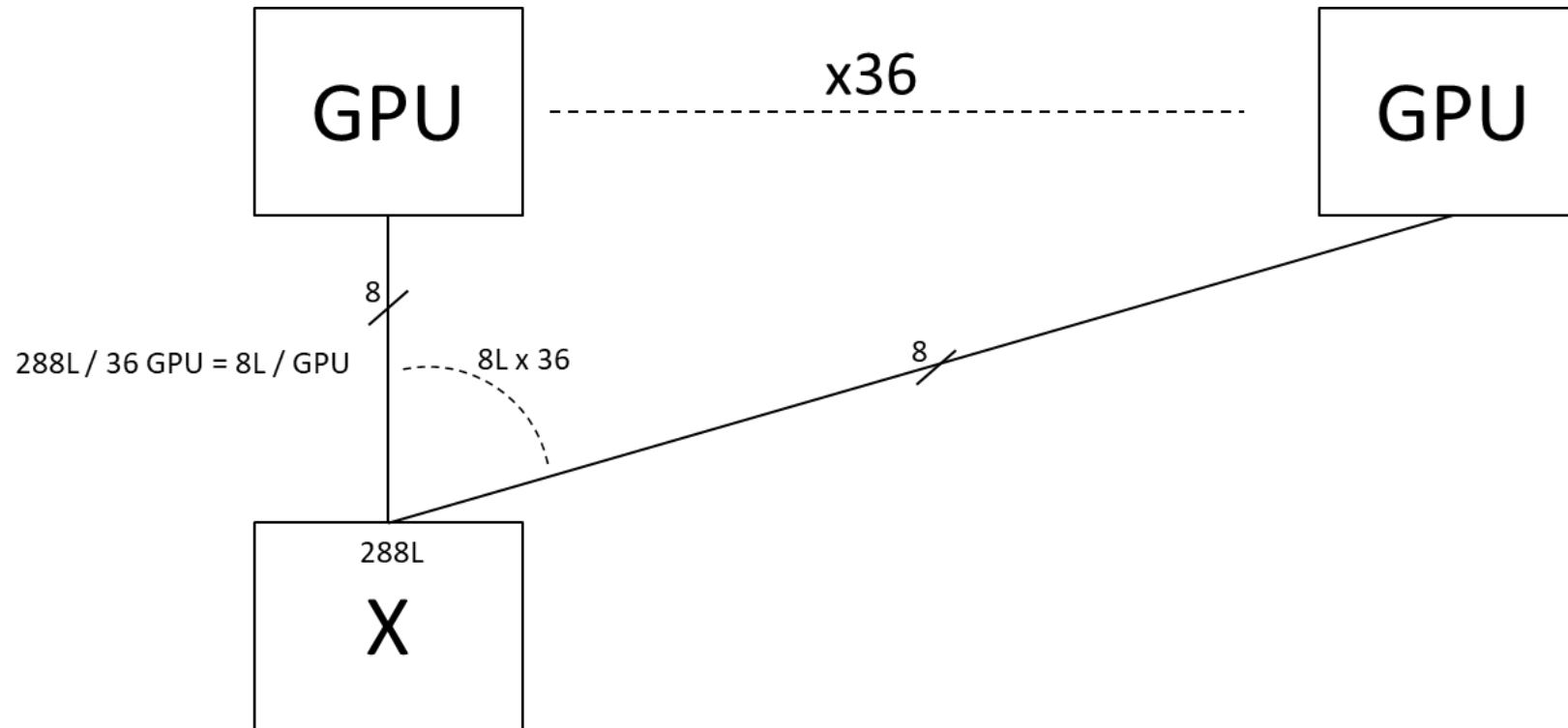
4. <https://docs.broadcom.com/doc/scale-up-ethernet-framework>

5. <https://ualinkconsortium.org/>

Scale-Up Topology

Data Center

L10 / L11



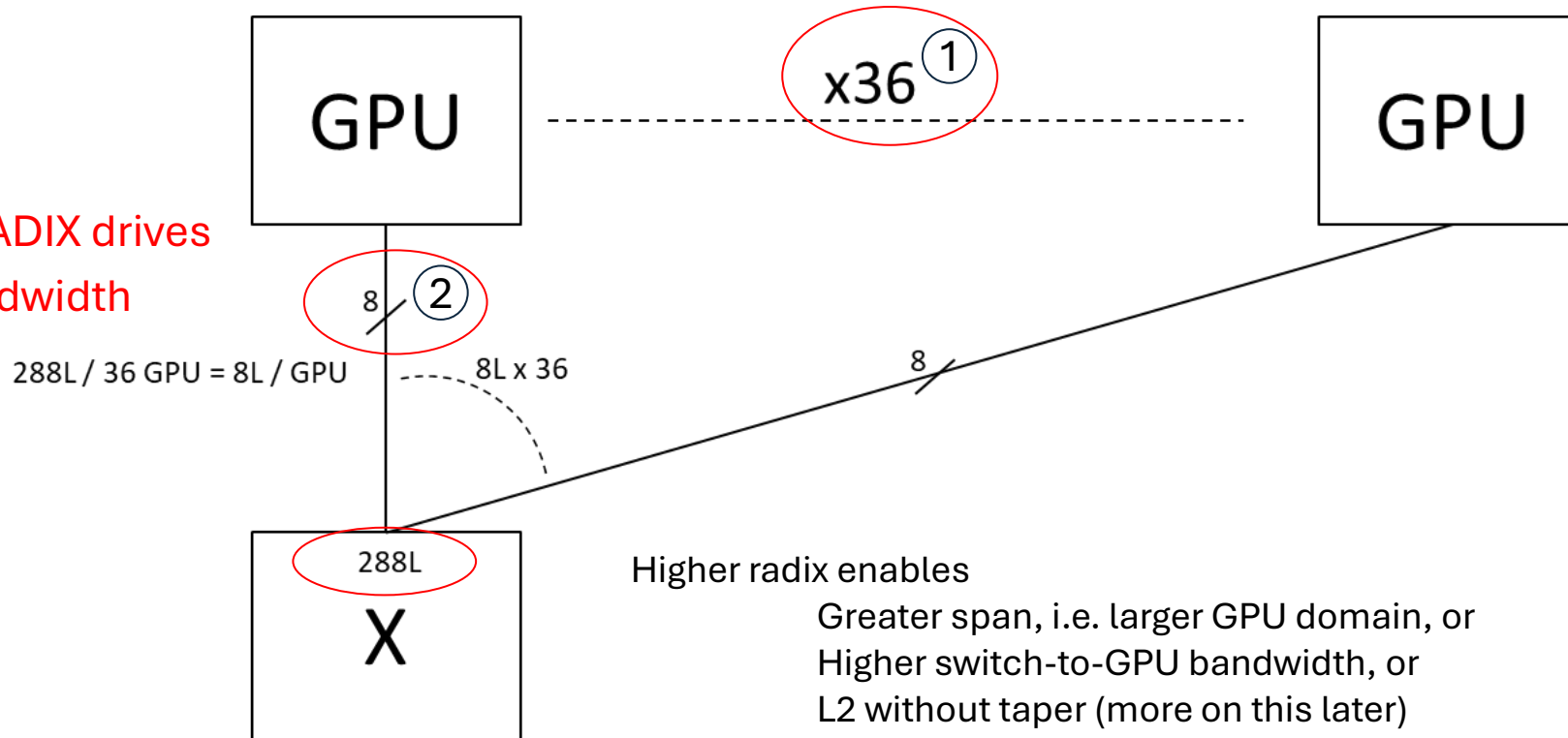
GPU Count and Switch Radix

Data Center

L10 / L11

1. SWITCH RADIX drives L1 domain span

2. SWITCH RADIX drives GPU-SW bandwidth

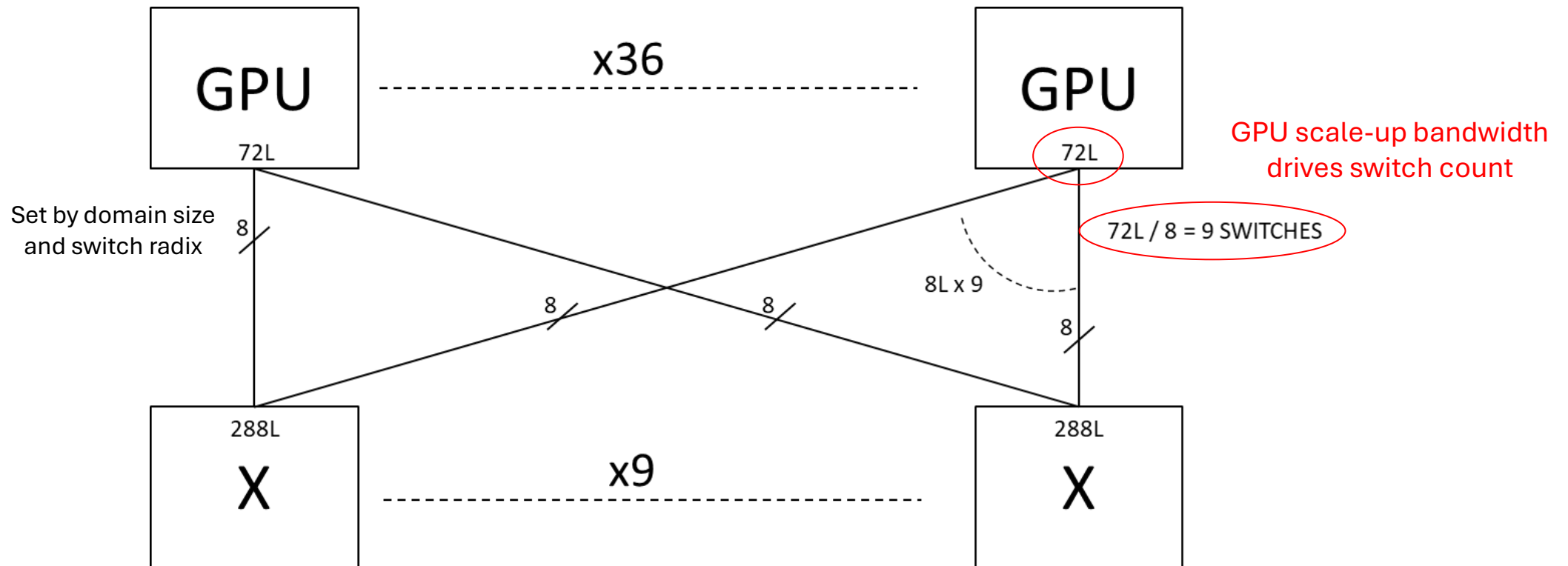


Higher radix enables
Greater span, i.e. larger GPU domain, or
Higher switch-to-GPU bandwidth, or
L2 without taper (more on this later)

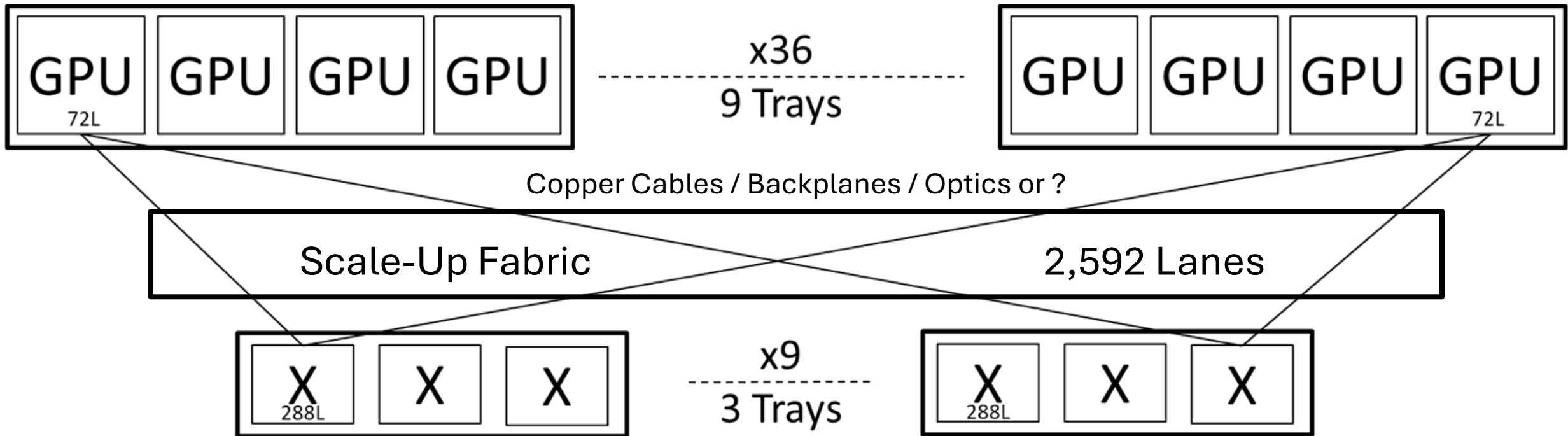
GPU Scale-Up Bandwidth and Switch Count

Data Center

L10 / L11



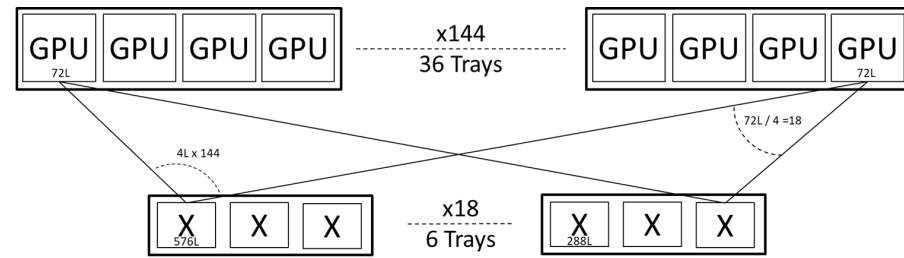
L10 / L11 Partitioning



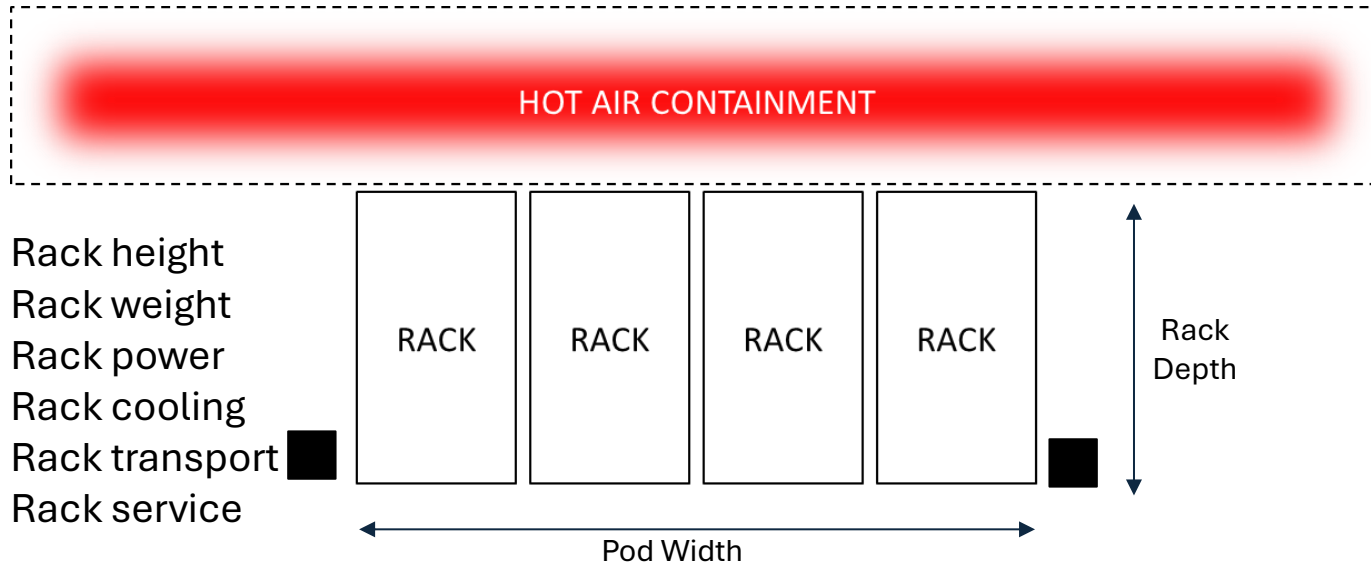
Emergent
Non-Obvious

How many lanes can exit switch and compute sleds?
SI from mechanical form factor?
Cable bulk in trays increase tray height?
Long paths need retimers: power, cost, heat → Not contributing to compute

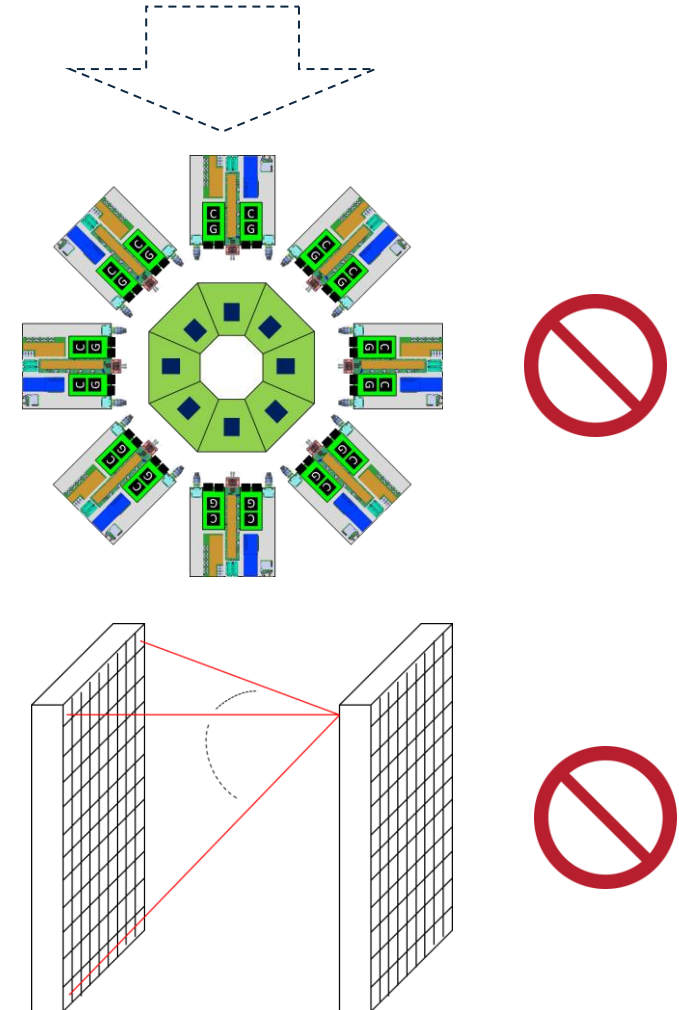
L10 / L11 Partitioning



TOP VIEW

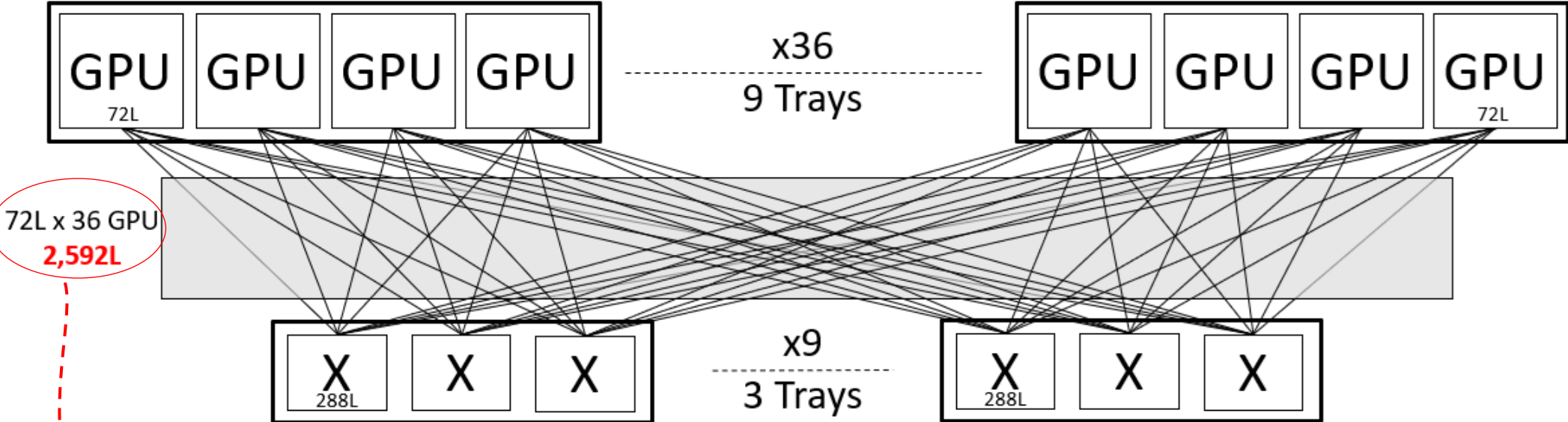


ROW DENSITY == \$\$\$\$



Scale-Up Fabric Complexity

*The **real** scale-up problem - fabric complexity*

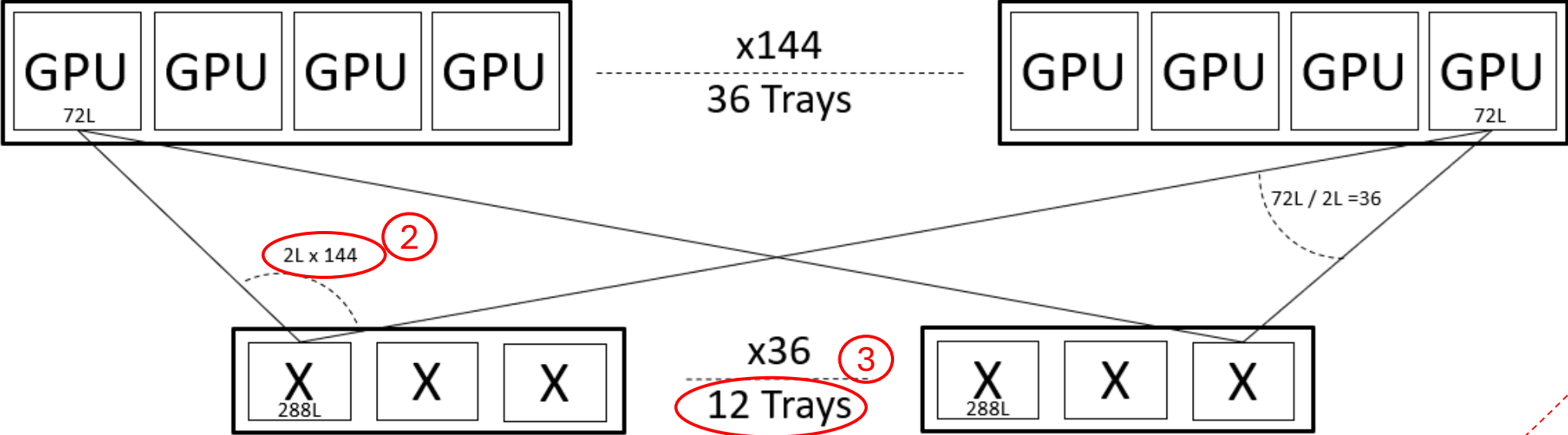


Now try scaling it to 144 GPU

44	COMPUTE SLED
43	COMPUTE SLED
42	COMPUTE SLED
41	COMPUTE SLED
40	COMPUTE SLED
39	COMPUTE SLED
38	COMPUTE SLED
37	COMPUTE SLED
36	COMPUTE SLED
35	
34	SWITCH SLED
33	SWITCH SLED
32	SWITCH SLED
31	
30	
29	
28	
27	
26	
25	
24	
23	
22	
21	
20	
19	
18	
17	
16	
15	
14	
13	
12	
11	
10	
9	
8	
7	
6	
5	
4	
3	
2	
1	

Scaling to Larger Domains

$144 \text{ GPU} \times 72\text{L} = 10,368\text{L}$ ①

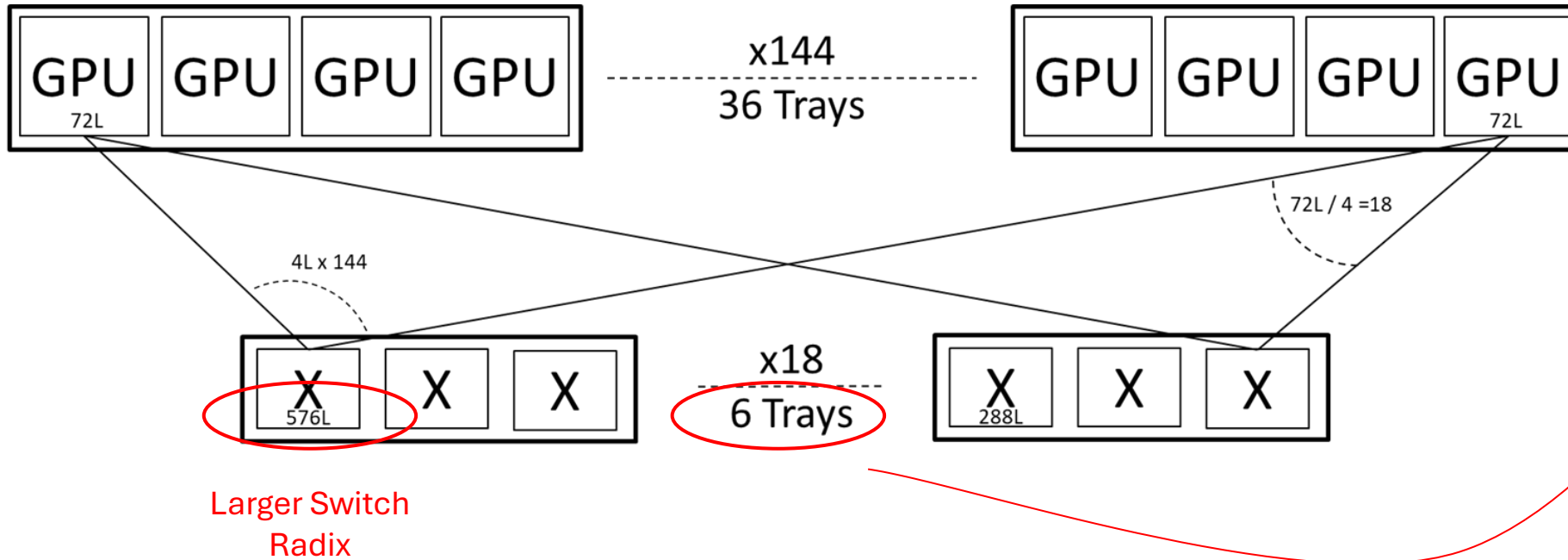


TOO MANY TRAYS

What if we have a larger switch?

44	COMPUTE SLED
43	COMPUTE SLED
42	COMPUTE SLED
41	COMPUTE SLED
40	COMPUTE SLED
39	COMPUTE SLED
38	COMPUTE SLED
37	COMPUTE SLED
36	COMPUTE SLED
35	COMPUTE SLED
34	COMPUTE SLED
33	COMPUTE SLED
32	COMPUTE SLED
31	COMPUTE SLED
30	COMPUTE SLED
29	COMPUTE SLED
28	COMPUTE SLED
27	COMPUTE SLED
26	SWITCH SLEDS 8 OU
25	
24	
23	
22	
21	
20	
19	
18	COMPUTE SLED
17	COMPUTE SLED
16	COMPUTE SLED
15	COMPUTE SLED
14	COMPUTE SLED
13	COMPUTE SLED
12	COMPUTE SLED
11	COMPUTE SLED
10	COMPUTE SLED
9	COMPUTE SLED
8	COMPUTE SLED
7	COMPUTE SLED
6	COMPUTE SLED
5	COMPUTE SLED
4	COMPUTE SLED
3	COMPUTE SLED
2	COMPUTE SLED
1	COMPUTE SLED

L1 Scale-Up - Larger Radix Switch



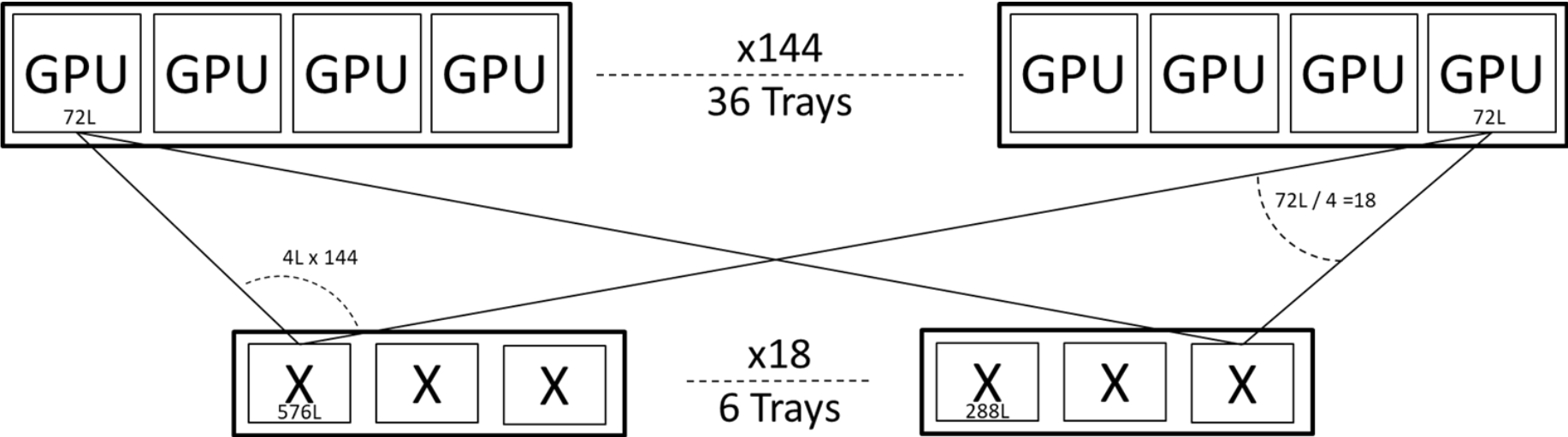
Not bad. Seems to fit

44	COMPUTE SLED
43	COMPUTE SLED
42	COMPUTE SLED
41	COMPUTE SLED
40	COMPUTE SLED
39	COMPUTE SLED
38	COMPUTE SLED
37	COMPUTE SLED
36	COMPUTE SLED
35	COMPUTE SLED
34	COMPUTE SLED
33	COMPUTE SLED
32	COMPUTE SLED
31	COMPUTE SLED
30	COMPUTE SLED
29	COMPUTE SLED
28	COMPUTE SLED
27	COMPUTE SLED
26	SWITCH SLEDS 8 OU
25	
24	
23	
22	
21	
20	
19	
18	COMPUTE SLED
17	COMPUTE SLED
16	COMPUTE SLED
15	COMPUTE SLED
14	COMPUTE SLED
13	COMPUTE SLED
12	COMPUTE SLED
11	COMPUTE SLED
10	COMPUTE SLED
9	COMPUTE SLED
8	COMPUTE SLED
7	COMPUTE SLED
6	COMPUTE SLED
5	COMPUTE SLED
4	COMPUTE SLED
3	COMPUTE SLED
2	COMPUTE SLED
1	COMPUTE SLED

X

Reality Check - Fabric Design Complexity

$18 \times 576L \times 2 \text{ DP/L} = 20,736 \text{ DP}$



44	COMPUTE SLED
43	COMPUTE SLED
42	COMPUTE SLED
41	COMPUTE SLED
40	COMPUTE SLED
39	COMPUTE SLED
38	COMPUTE SLED
37	COMPUTE SLED
36	COMPUTE SLED
35	COMPUTE SLED
34	COMPUTE SLED
33	COMPUTE SLED
32	COMPUTE SLED
31	COMPUTE SLED
30	COMPUTE SLED

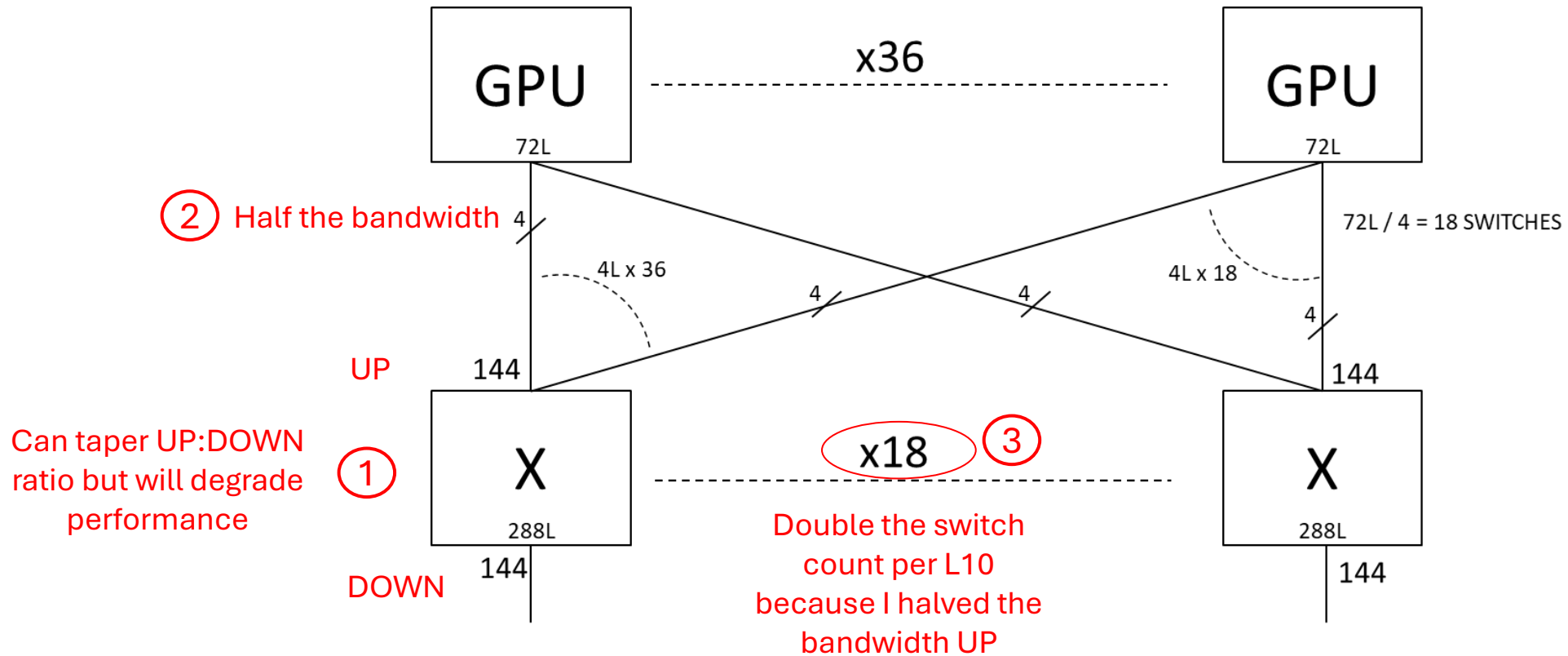
- Mechanical design
- Lanes per sled
 - Connectors
 - Flyovers
 - PCB losses
 - Manufacturability
 - Throughput
 - Yield
 - MTBF
 - Size and Weight
 - Shipping methods
 - Shock and Vibe
 - Data Center Installation
 - Serviceability
 - Power (if re-timed)
 - Cooling (If re-timed)

What if I don't have a larger switch?

Implications of L2 Scale-Up

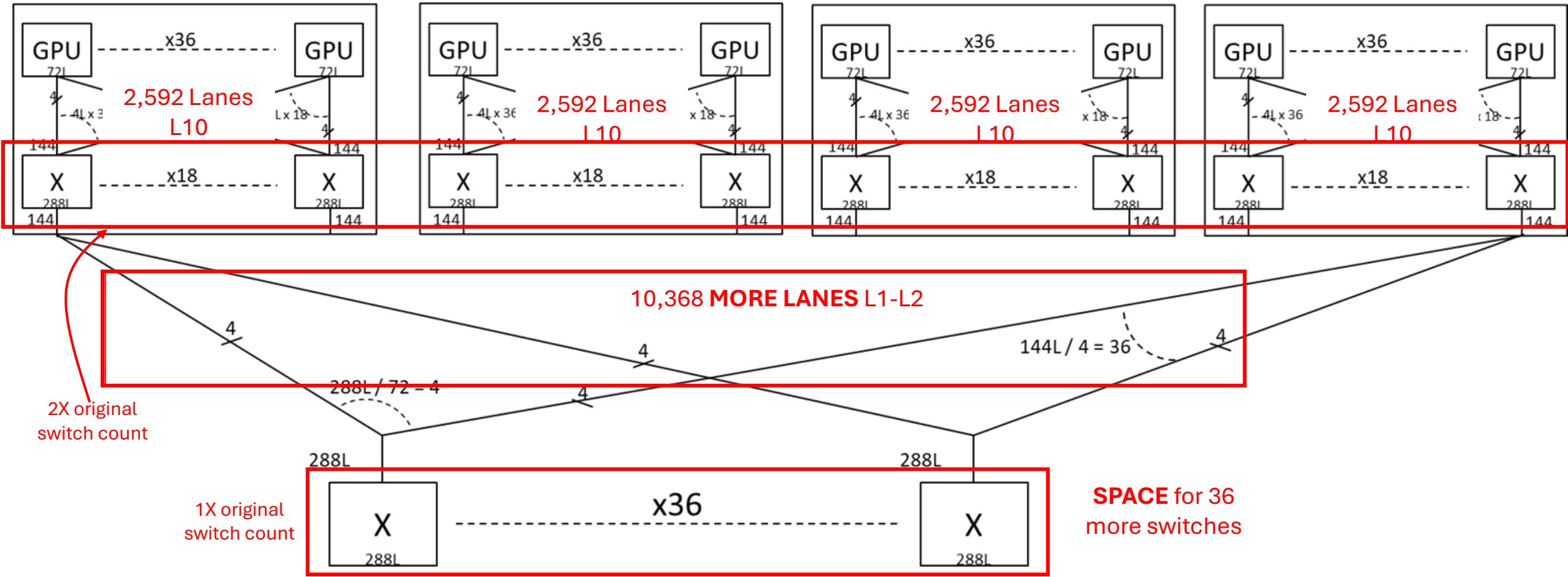
Data Center

L10 / L11



Implications of L2 Scale-Up (It's not great)

10,368L GPU to L1

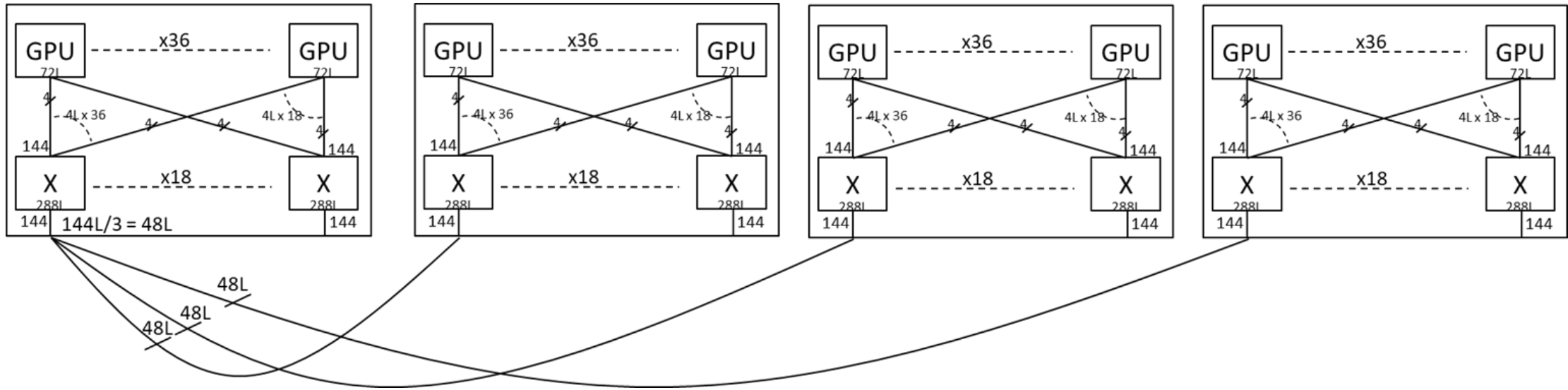


3X switch count of L1 (Space, Power, Cooling, Retimers, SI)

Alternate Topology “L1.5”

Data Center

L10 / L11



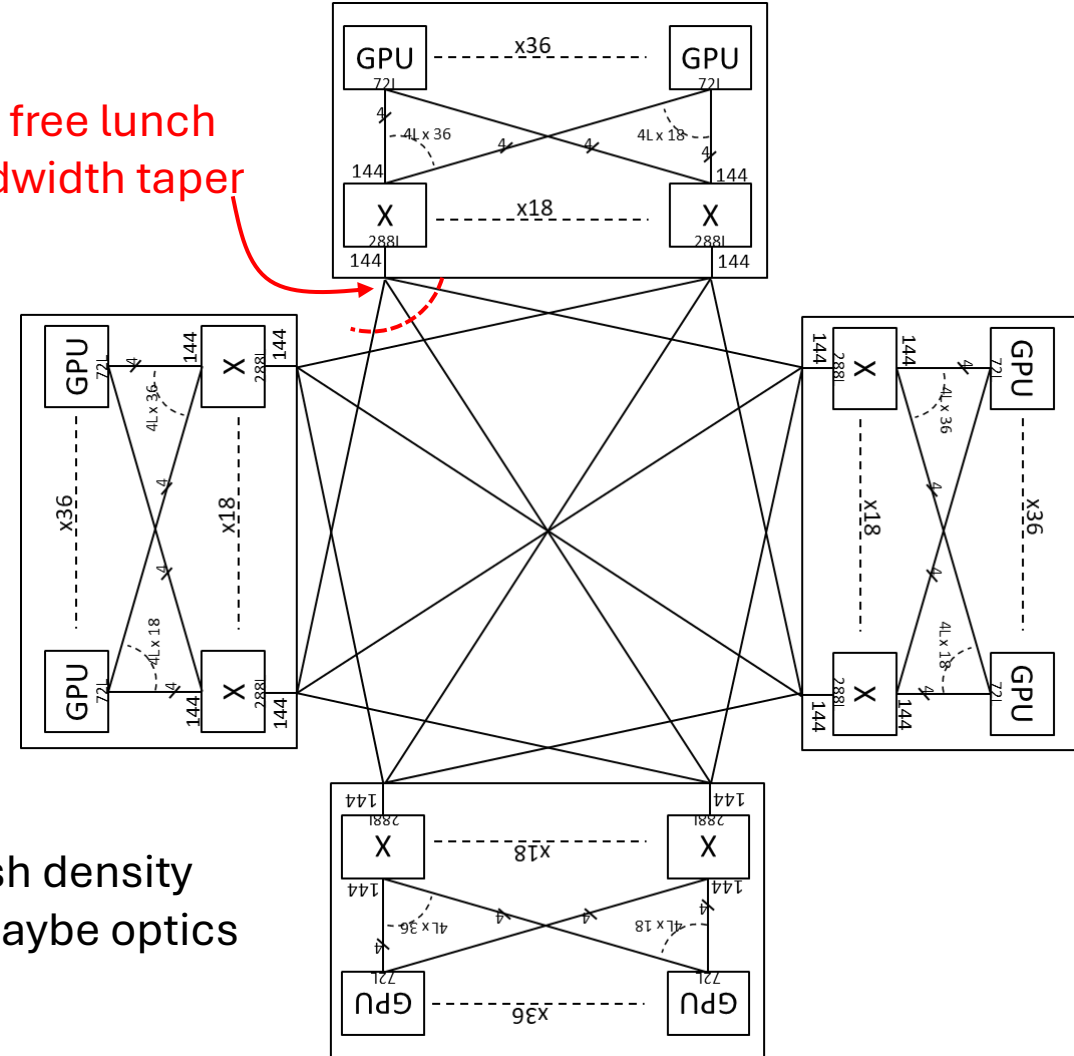
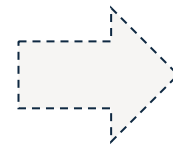
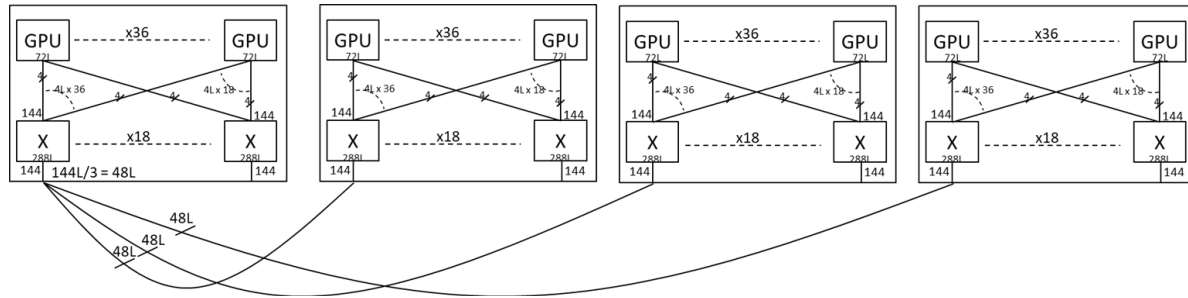
L1.5 is a “Mesh of Domains”

Decent compromise for scale-up

It's really this

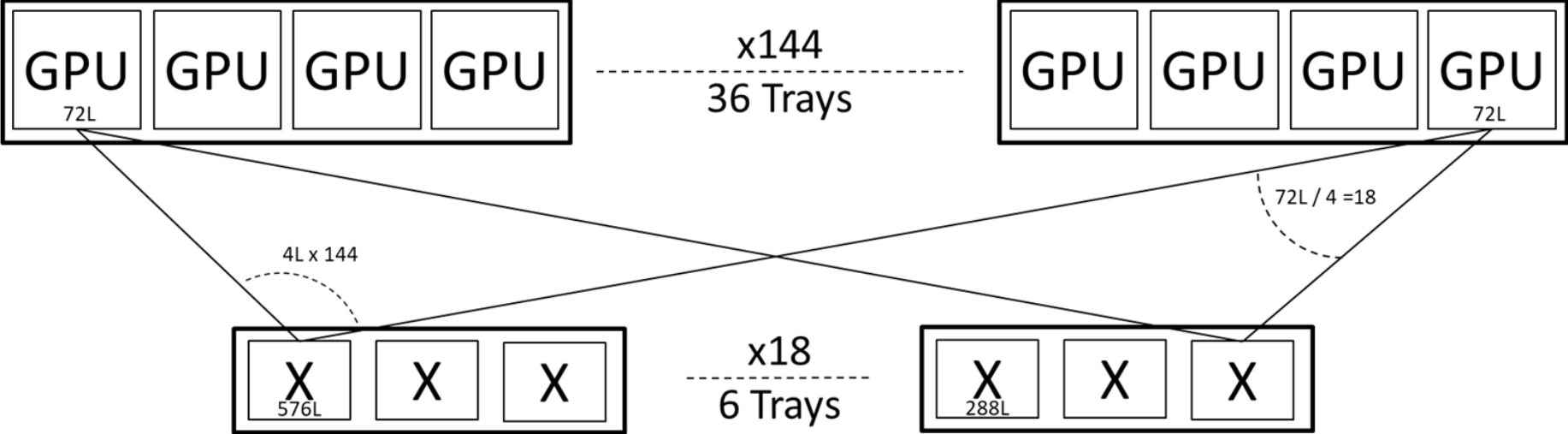
Stare at L1.5 long enough

Still no free lunch
3:1 bandwidth taper



Reasonable mesh density
Use copper or maybe optics

Scale-Up Fabric Signal Integrity



26 OU x 48mm = 1.25 m + ?
 Not terrible, but what about:
 Mechanical partitioning, cable considerations, connectors, PCB, vias... ?

44	COMPUTE SLED
43	COMPUTE SLED
42	COMPUTE SLED
41	COMPUTE SLED
40	COMPUTE SLED
39	COMPUTE SLED
38	COMPUTE SLED
37	COMPUTE SLED
36	COMPUTE SLED
35	COMPUTE SLED
34	COMPUTE SLED
33	COMPUTE SLED
32	COMPUTE SLED
31	COMPUTE SLED
30	COMPUTE SLED
29	COMPUTE SLED
28	COMPUTE SLED
27	COMPUTE SLED
26	SWITCH SLEDS 8 OU
25	
24	
23	
22	
21	
20	
19	
18	COMPUTE SLED
17	COMPUTE SLED
16	COMPUTE SLED
15	COMPUTE SLED
14	COMPUTE SLED
13	COMPUTE SLED
12	COMPUTE SLED
11	COMPUTE SLED
10	COMPUTE SLED
9	COMPUTE SLED
8	COMPUTE SLED
7	COMPUTE SLED
6	COMPUTE SLED
5	COMPUTE SLED
4	COMPUTE SLED
3	COMPUTE SLED
2	COMPUTE SLED
1	COMPUTE SLED

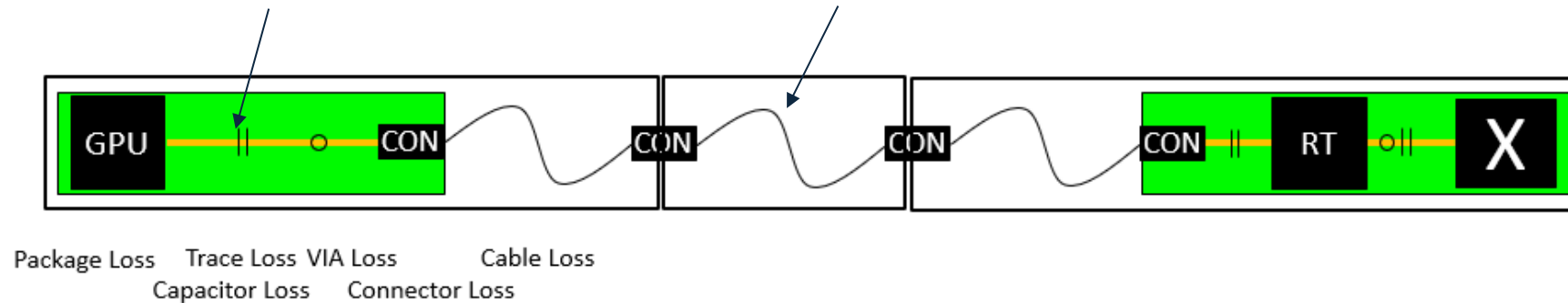
Scale-Up Channel

- PCB Insertion loss 1.27 dB/in in PCB for 212G PAM4¹, 1.4 dB/in for 224G PAM4², 1.5 dB per via
- Copper TwinAx loss 0.19 - 0.34 dB/in depending on AWG
- Connectors Varies. < 1.0 to 10s of dB
- Optical 50 um fiber loss³ @ 850 nm is 76×10^{-6} dB/in (3 dB/km)

AWG	dB/m	dB/in
26	7.625	0.19
27	8.875	0.23
30	11.6	0.29
32	13.4	0.34

12 in of PCB loss = 15.24 dB

2m of 27 AWG TwinAx loss = 17.75 dB



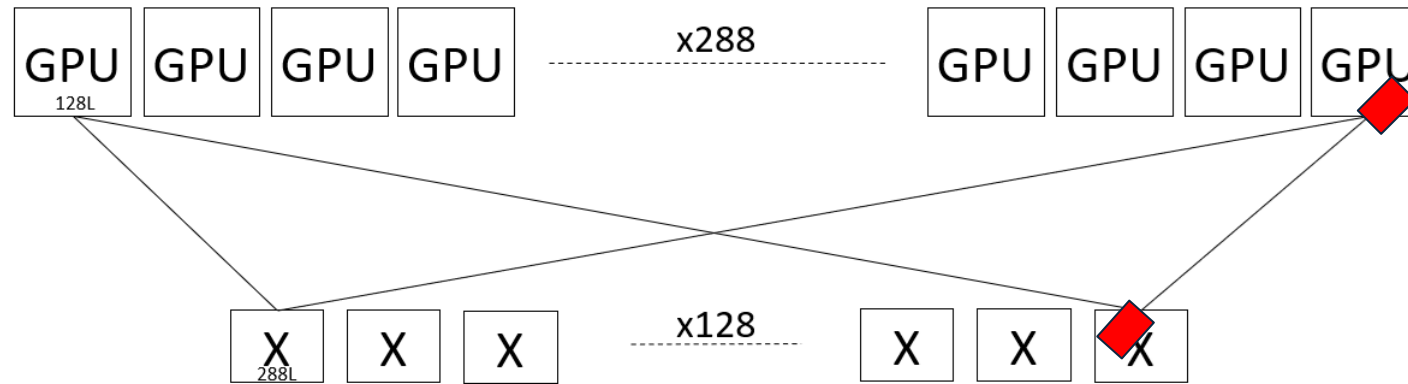
1. "A 212.5 Gbps-PAM4 1 Meter DAC Long Reach Channel and Its Characteristics" https://www.ieee802.org/3/dj/public/23_09/lim_3dj_03_2309.pdf
2. "Next-Generation PCB Loss Analysis" <https://www.signalintegrityjournal.com/articles/3159-next-generation-pcb-loss-analysis>
3. 50 um fiber loss, 850 nm multi-mode 3 dB/km, 1300 nm multi-mode 0.7 dB/km, 39,370 in/km

So why not pluggable optical SFPs for scale-up?

Hypothetical 288 GPU cluster

36,864 lanes for scale-up

212G per lane requires 4,608 x 3.2 Tbps (16L) OSFP



Capex @ \$4K per OSFP
MTBF for 4,608 OSFP²

\$18 M
9 days

“We’re going to need better optics”

1. 288 GPU x 128L = 36,384 lanes / (16 x 212G Lanes per OSFP-XD) x 2 OSFP per MPO = 4,608 OSFP
2. Assuming OSFP MTBF 1M hours³, 4,608 x OSFP = 1,843 FPMH = 543 h MTBF
3. MTBF passive copper cable 500 M hours : https://www.etulinktechnology.com/blog/dac-and-aoc-who-will-be-the-winner-in-the-field-of-data-communication-_b317
MTBF active copper cable 50 M hours : https://www.etulinktechnology.com/blog/dac-and-aoc-who-will-be-the-winner-in-the-field-of-data-communication-_b317
MTBF active optical cable 1 M hours : https://stordis.com/wp-content/uploads/documents/Presentation_Webinar-STORDIS-CREDO_2022-04-21.pdf

Closing Thoughts on Scale-Up

Directly impacts model **training** time and **performance**

Complexity driven by **GPU count**, **scale-up BW** and **switch radix**

Bounded by **datacenter**, **mechanical** and **SI** requirements

Will benefit from new tech such as >212G, optics, etc.

